Cracking the Energy Industry's Data Dilemma

The Evolution of Database Technology is Unlocking the Potential of Real-Time Data

The oil industry is awash with data. Companies collect information from a multitude of sensors attached to equipment on every rig, from the drilling floor, from pressure tests, and even from the drill bit as it chews its way through rock thousands of feet below the surface.

Data pours in from well sites, fracking equipment, pipelines and gas gathering systems. A company's fleet of frack trucks, lined up at the well site for a typical pressure pumping job, might transmit data to headquarters over hundreds and possibly thousands of channels a second, reporting on everything from maintenance needs for the truck itself to real-time pressure data for the reservoir engineer.

Most companies struggle just to collect, store and analyze all this information. Yet the lack of real-time responsiveness limits the benefits of all this data. The competitive advantage for companies operating in the oilfields of the future will depend on their ability to process vast amounts of data on a large scale, and provide real-time actionable intelligence to remote locations that will improve operating efficiency, safety and maintenance.

A Deluge of Data

In the past five years, the amount of data generated by oil and gas operations has increased dramatically — there's more of it, and it's arriving more quickly. Companies are finding that the volumes of data they gather are beginning to stretch the physical limits of the equipment on which they collect and store it.

In years past, the oil and gas industry's biggest challenge in dealing with real-time data was transmission. The remote locations of well sites, both onshore and offshore, as well as the vast expanse of pipelines, made it difficult to connect sensors in the field to a central data center. In some cases, data collected on an offshore rig might be discussed with engineers in the base office via radio each morning, then stored on a compact disc or thumb drive and physically carried by crew members on the next helicopter flight to shore.

Once this data arrived, it was fed into a single server large enough to store and analyze all the information, then the results of the analysis were distributed back to the field in the same manner in which the data arrived.

Over the past decade, energy companies have increased fiber networks and other remote communications capabilities. Many deepwater rigs in the Gulf of Mexico, for example, are now networked with high-speed fiber optics.

As communication has improved, the challenge has passed from collecting data to managing and analyzing it. Today, cloud computing and streaming data are transforming the database industry, and the implications for the energy industry are profound for safety, maintenance, efficiency and cost-effectiveness. Many companies, however, have struggled to keep up with the technological advances. As a result, they often do not have access to the full picture of remote operations. Data on equipment installed on certain wells, or a comparison of drilling data may not be readily available.

Database Evolution

In the 1980s, most energy companies used centralized data management systems that stored all data on a single mainframe, usually at or near the company's headquarters. In the 1990s, data management evolved to the client-server architecture running relational databases, which allowed for more sharing of data, but still required centralized storage.

These systems no longer can keep pace with today's radically connected environment. Companies need databases that are flexible and scalable, without restrictions based on the location of inputs or data storage. In other words, they must be able to write and read data from anywhere at any time, and they need the data synchronized across their networks instantly.

The relational databases of the 1990s are fundamentally ill-equipped to cope with these demands, and as a result, many oil and gas companies are not benefitting from the investment they made in data management.

As the capabilities of big data and the "Internet of Things" continue to expand, companies whose system architecture is based on relational database software will find it increasingly difficult to maintain a competitive advantage.

Companies that want to capture the full benefit of real-time data need systems that require less maintenance and less intervention, with built-in redundancies to protect against data losses. They also want a system that is reliable and unaffected by outages in a single location, another vital characteristic.

As a result, many companies need to rethink their approach to real-time data from the bottom up.

Distributed Architecture

Instead of continuing to build on an existing relational architecture, companies that want to leverage the full potential of real-time data should adopt a flexible, scalable architecture, in which clusters of servers all share images of the same data that can be accessed from anywhere.

Data management for this new era requires four key capabilities:

- Underlying architecture. For a company to service multiple independent locations globally, it needs database clusters distributed worldwide, all of which have the same data. This allows access of all data from any location. In simple terms, it is the same technology that allows someone to start a streaming video on a device in one location and finish it on another device somewhere else.
- Scale and performance. As the flow of data intensifies, companies must know they can manage more of it, and from more users, without overloading the system. A ddistributed architecture enables the seamless addition of more capacity without downtime
- Continuous availability. Because the data is replicated and stored in multiple locations, the system adjusts to an outage in one area by simply routing traffic to another, avoiding downtime.
- Manageability. The software must be simple enough that personnel can easily manage and maintain the system without extensive software expertise.

These characteristics create a data management network that responds to a company's needs at high speed. Large amounts of data can be assimilated quickly and analyzed in real time to help make better business decisions. The platform supports a mixed workload environment that can run transactional, analytical, and search functions simultaneously on the same data in the same database.

Applications in the Field

Applied to oil and gas operations, a distributed architecture enables companies to analyze data and deliver it back to remote interested parties almost instantaneously, dramatically reducing lost time. Mud pumps, for example, can be

monitored for efficiency and send alerts of internal parts wear or other issues, allowing for proactive maintenance and reducing costly and unnecessary shutdowns.
Real-time data, however, can do more than just improve machine efficiency and reduce maintenance costs. It also can enable drilling engineers to determine when they are approaching overpressure zones and take steps, such as adjusting hydrostatic pressure in the well bore, to ensure that drilling operations proceed correctly. This can help improve efficiency in the drilling process and improve safety.

Real-time server clusters can collect a litany of details from the wellbore itself – the drill bit's revolutions per minute, the weight on the bit, the rate of penetration per hour, mud pumping pressure and volume – and alert drilling engineers to any discrepancies. The ability to analyze drilling parameters in real time dramatically increases operational and safety management.

More important, it ensures that all well management personnel have access to the same information at the same time, which fosters better coordination of decision making. Operators can also see how one well or rig is performing compared with another, which can further improve operating efficiency and provide a more comprehensive and comparative risk assessment of each location.

The technology can benefit pipeline operators as well. They have been sending smart pigs through their pipeline networks for years, collecting data on corrosion and the thickness of pipe walls. But extracting and processing that data used to take weeks. Now, the data can be captured an analyzed much more efficiently, enabling pipeline operators to address potential maintenance issues more quickly and reducing the chance of spills and leaks.

The Database of the Future

The greatest advantage of distributed architecture is that it was tailor-made for cloud computing. As the demands on databases continue to grow, systems will require even greater flexibility. For example, a single application may have different components, customer profiles, catalogs and other types of data, each of which has different needs for managing and storing the data in the same database.

During a time of weak commodity prices, energy companies may be reluctant to invest in a new software platform, especially if they already have a system that they are familiar with.
DataStax has built its software on Cassandra, the leading open-source database software. With Cassandra as the core, Datastax Enterprise, is not only more cost effective, it is also more flexible and scalable. The distributed architecture enables customers to handle disparate types of data with an "all in one" approach that can quickly expand to meet increasing capacity.

Systems built on DataStax Enterprise can also handle the industry's common remote collection and sharing requirements. Facilitating these remote application needs in low bandwidth or outage situations solves a persistently challenging problem while also delivering bandwidth efficiencies in difficult situations. This flexibility also enables companies to use the same data models and applications in the field and at the data center or cluster.

Databases of the future must be more scalable and flexible, to manage and analyze not only more and different types of data, but to increase the speed with which business and operating decisions are made. In the oil and gas industry, small windows of opportunity can drive big results. Having access to the data in real time can make the difference in capitalizing on those opportunities.